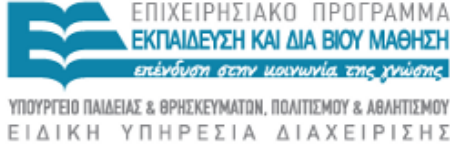




Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

ΠΑΡΑΔΟΤΕΟ 7.1

Έκθεση Ελέγχου της Βάσης Δεδομένων

ΠΑΚΕΤΟ ΕΡΓΑΣΙΑΣ 7

«ΘΑΛΗΣ-ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ-ΔΙΑΦΩΝΗΝ: φωνήεντα των ελληνικών ΔΙΑΛΕΚΤΩΝ: Φωνολογική και φωνητική ανάλυση, κατάρτιση διαλεκτικού ευρετηρίου και κοινωνική Ευαισθητοποίηση για την αΝάδειξη του γλωσσικού πλούτου»

ΔΙΑΦΩΝΗΝ



VOCALECT

MIS: 379396

ΥΠΕΥΘΥΝΗ ΟΜΑΔΑ: ΕΟ3



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα

1. Εισαγωγή	3
2. Έλεγχος λειτουργικών προδιαγραφών	4
3. Έλεγχος του λογισμικού που αναπτύχθηκε	6
4. Συμπεράσματα	11



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

1.Εισαγωγή

Η σύνθετη διαδικτυακή βάση δεδομένων δημιουργήθηκε με γνώμονα: πρώτον, την οργάνωση και παρουσίαση των αποτελεσμάτων, ώστε να προσφέρονται για τη μελέτη και άλλων επιπέδων γλωσσολογικής ανάλυσης (π.χ. συντακτικής, μορφολογικής) και δεύτερον, τη δυνατότητα διαρκούς εμπλουτισμού της με υλικό από μελλοντικές μελέτες. Στη βάση αποθηκεύονται τα γλωσσικά δεδομένα τα οποία συνδέονται με γεωγραφική και μεταγλωσσική πληροφορία. Με αυτόν τον τρόπο συνδυάζονται γλωσσολογικά δεδομένα και γεωγραφική ποικιλία και καθίσταται εφικτή η δημιουργία σύνθετων αναζητήσεων βάσει γλωσσολογικών και μεταγλωσσικών κριτηρίων. Η σύνθετη διαδικτυακή βάση παρέχει τη δυνατότητα πρόσβασης στα δεδομένα που θα συλλεχθούν. Ο ρόλος που κατέχει στο γενικό πλαίσιο του έργου είναι αυτός του πυρήνα της διαδικασίας έκθεσης των αποτελεσμάτων, αφού περιέχει την πληροφορία στο σύνολό της που εφόσον συνδυαστεί με τον κατάλληλο τρόπο παρέχει μια λεπτομερή σύγκριση μεταξύ των χαρακτηριστικών των υπό μελέτη ελληνικών διαλέκτων.

Ο συνδυασμός της διαλεκτικής βάσης δεδομένων και του σύνθετου ευρετηρίου θα πραγματοποιηθεί με τη βοήθεια ενός δυναμικού ιστοχώρου. Ο σχεδιασμός του θα πρέπει να είναι τέτοιος που θα εξυπηρετεί τη συνεργασία των ερευνητικών ομάδων, καθώς και την παρουσίαση των ερευνητικών αποτελεσμάτων μέσα από την υλοποίηση των διεπαφών για το σύνθετο ευρετήριο.

Στα πλαίσια του παραδοτέου αυτού έγινε έλεγχος της λειτουργίας της Βάσης Δεδομένων και εισηγήθηκαν μικρές διορθώσεις, οι οποίες και υλοποιήθηκαν.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
Πρόγραμμα για τη ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

2. Έλεγχος λειτουργικών προδιαγραφών

Η βάση δεδομένων που έχει αναπτυχθεί αποθηκεύει όλη την πληροφορία που χρειάζεται ο ιστοχώρος για να λειτουργήσει. Τα βασικά δεδομένα είναι αρχεία ήχου, αρχεία κειμένου πρωτογενή δεδομένα ηχογραφημένου λόγου, επεξεργασμένα δεδομένα, μεταδεδομένα τα οποία θα χαρακτηρίζουν τα ηχητικά αρχεία της και τέλος όλα τα αρχεία κώδικα για την ομαλή λειτουργία της. Με γνώμονα τον όγκο των δεδομένων η βάση σχεδιάστηκε ώστε να παρέχει άμεση πρόσβαση στα δεδομένα της για τα μέλη των ερευνητικών ομάδων που συμμετέχουν στο πρόγραμμα και ταυτόχρονα αποδοτική, όσον αφορά την ταχύτητα, σύνδεση και επικοινωνία με αυτή.

Βασικές προϋποθέσεις για το σύστημα που προνοεί για την αποθήκευση της προαναφερθείσας πληροφορίας είναι η ασφάλεια των δεδομένων και του συστήματος, η γρήγορη απόκριση και επικοινωνία με τον ιστοχώρο και μεγιστοποίηση του διαθέσιμου χώρου. Για την διασφάλιση των παραπάνω προϋποθέσεων η βάση υλοποιήθηκε σε έναν ιδιόκτητου server. Τα αρχεία αποθηκεύονται στο σκληρό δίσκο του server και οι πληροφορίες που τα συνοδεύουν καθώς και η διαδρομή (path) αποθηκεύονται στο σύστημα της βάσης. Ο διαθέσιμος χώρος αποθήκευσης είναι εύκολα επεκτάσιμος προσθέτοντας επιπλέον σκληρούς δίσκους στον υπολογιστή, έτσι δεν υπάρχει πλέον περιορισμός στον όγκο δεδομένων που θα μπορεί να φιλοξενήσει η βάση δεδομένων. Επίσης, τα δεδομένα αποθηκεύονται σε ένα κοινό μέσο στο οποίο ο διαχειριστής είναι ταυτόχρονα και μέλος της ερευνητικής διαδικασίας εξαλείφοντας έτσι την πιθανότητα έκθεσης των προσωπικών δεδομένων των ομιλητών. Τέλος, για τη μετάδοση των δεδομένων η επικοινωνία πραγματοποιείται μεταξύ ενός τερματικού σταθμού και μόνο ενός μηχανήματος (αυτό του server) αυξάνοντας έτσι την απόδοση χωρίς να υπάρχουν καθυστερήσεις από επικοινωνία δια μέσου διαφορετικών υποδικτύων.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

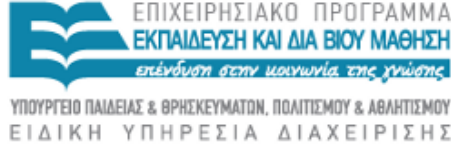
Για την εξυπηρέτηση της βασικής απαίτησης για την υλοποίηση της βάσης δεδομένων, όσον αφορά στη λειτουργικότητά της αλλά και την ευελιξία σε μελλοντικές αλλαγές που τυχόν πραγματοποιηθούν, επιλέχθηκε η MySQL σε συνδυασμό με το phpMyAdmin ως το καταλληλότερο Σύστημα Διαχείρισης Βάσης Δεδομένων (DBMS). Αυτά δίνουν τη δυνατότητα για υψηλού επιπέδου λειτουργίες όπως τον ορισμό, τη δημιουργία, την αναζήτηση, την ενημέρωση, και τη διαχείριση της βάσης. Τα εργαλεία αυτά καθιστούν ξεκάθαρη την υλοποίησή της βάσης και έτσι η σύνδεση με αυτή με τον ιστοχώρο και τον διαχειριστή είναι πολύ εύκολη και γρήγορη διαδικασία. Με αυτόν τον τρόπο, τα δεδομένα που αποθηκεύονται σε αυτή μπορούν ακόμη και να μεταφερθούν σε μια διαφορετική βάση δεδομένων. Πιθανώς για τη μεταφορά, μπορεί τα δεδομένα να χρειαστεί να κωδικοποιηθούν ώστε να ακολουθούν κάποιο συγκεκριμένο πρότυπο κάτι που δεν αποτελεί εμπόδιο στη διαδικασία.

Η πρόσβαση της βάσης εμπεριέχει επεξεργασία των δεδομένων κάνοντας χρήση του DBMS και αναζήτηση και παράθεσή τους μέσα από ένα σύνθετο ευρετήριο. Πραγματοποιείται μέσα από έναν ιστοχώρο ο οποίος λειτουργεί με το συνήθη τρόπο και το μόνο που θα απαιτεί είναι πρόσβαση στο internet. Ο συνδυασμός των παραπάνω τεχνολογιών υποστηρίζει την ταυτόχρονη εξυπηρέτησή πολλών χρηστών από τη βάση.

Έγινε έλεγχος των λειτουργικών προδιαγραφών της Βάσης Δεδομένων και έγινε βελτιστοποίηση με στόχο την βελτίωση της αποδοτικότητας της βάσης και τη μείωση του χρόνου απόκρισης.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

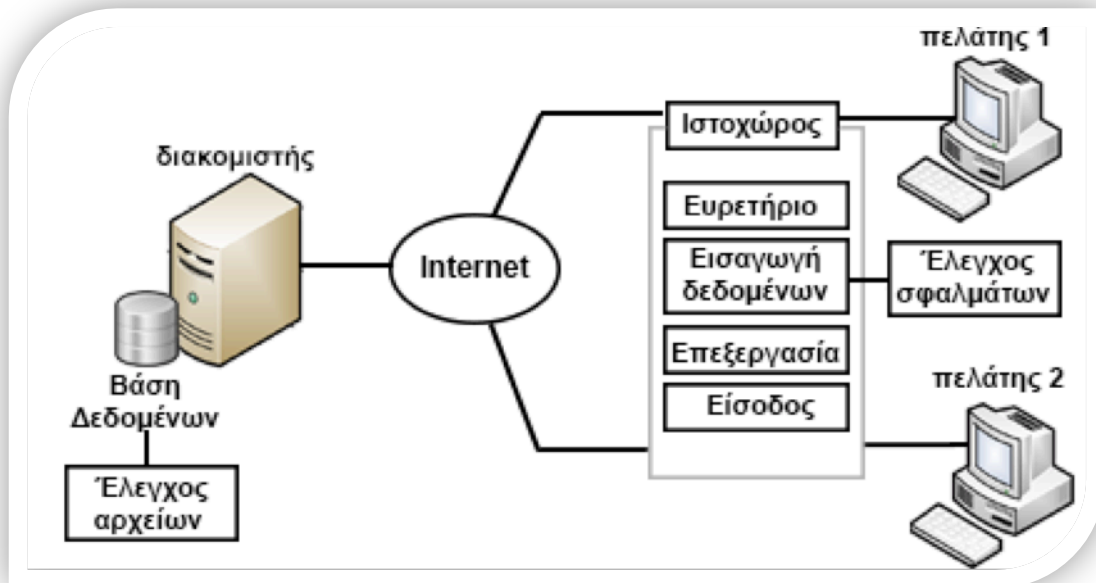


ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

3. Έλεγχος του λογισμικού που αναπτύχθηκε

Η διαδικασία που ακολουθήθηκε για την δημιουργία της βάσης και της επικοινωνίας μεταξύ του ιστοχώρου και αυτής είναι απλή και η πιο συνηθισμένη σε παρόμοια συστήματα. Συγκεκριμένα, χρησιμοποιήθηκε το γνωστό μοντέλο client-server (διακομιστή-πελάτη) το οποίο επιτρέπει την τμηματική υλοποίηση του συνόλου των λειτουργιών και την εκτέλεση κάθε μιας από αυτές στο κατάλληλο μηχάνημα. Η βάση δεδομένων υλοποιήθηκε εξ' ολοκλήρου στο μηχάνημα του server, το οποίο διαθέτει ένα πολύ μεγάλο χώρο αποθήκευσης και έτσι δεν υπάρχει ο κίνδυνος για περιορισμό στο μέγεθος και τον αριθμό των αρχείων που θα προέρθουν από τις διαδικασίες ηχογράφησης και φωνητικής μεταγραφής. Για τη διαχείρισή των πινάκων της παρέχεται ένα υποσύστημα, το phpMyAdmin, στο οποίο έχει πρόσβαση μόνο ο Διαχειριστής, από το οποίο μπορεί απλά να τροποποιήσει ή ακόμη να αλλάξει τη δομή της βάσης.

Όπως φαίνεται στο Σχήμα 1, στο οποίο απεικονίζεται το μοντέλο που ακολουθήθηκε



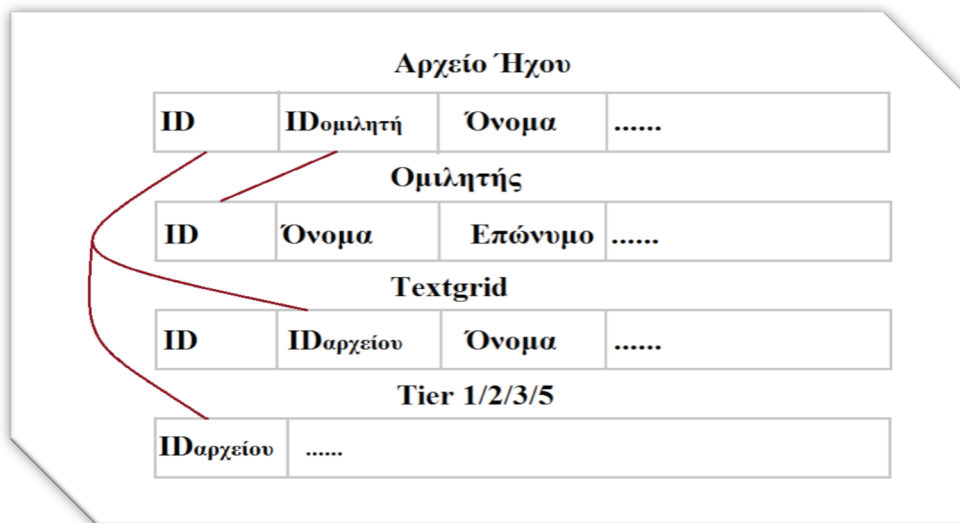
Σχήμα 2: Απεικόνιση του αρχιτεκτονικού μοντέλου

για τη σχεδίαση, όλες οι λειτουργίες που αφορούν τη διαλεκτική βάση πραγματοποιούνται μέσα από τον ιστοχώρο, τα αρχεία του οποίου είναι αποθηκευμένα στο server. Ένας τερματικός σταθμός συνδέεται σε αυτόν μέσω του διαδικτύου κάνοντας μια αίτηση για την ανάγνωση των αρχείων που αφορούν τις σελίδες του ιστοχώρου. Εύκολα προκύπτει το συμπέρασμα ότι κάθε αίτηση προς τη βάση δεδομένων εξυπηρετείται αφού εγκατασταθεί μια σύνδεση του πελάτη με αυτή μέσω της διεπαφής του χρήστη. Ανεξάρτητα του ρόλου δηλαδή που κατέχει κάθε χρήστης η προσπέλαση της βάσης ξεκινά και τελειώνει στην πλευρά του τερματικού σταθμού. Όλες οι δυνατότητες που παρέχονται στους απλούς χρήστες αλλά και στους ερευνητές του έργου, συνδέονται άμεσα με τη διαλεκτική βάση. Εξετάζοντας το υποσύστημα εισαγωγής δεδομένων, ένας ερευνητής επιλέγει κάποιο αρχείο που βρίσκεται αποθηκευμένο τοπικά στο μηχάνημά του ώστε να αποθηκευτεί στην βάση του απομακρυσμένου server. Πριν την τελική εισαγωγή γίνονται δύο ειδών έλεγχοι, ένας στην πλευρά του πελάτη και ένας στην πλευρά του server. Ο αρχικός έλεγχος αφορά την ορθή συμπλήρωση των μεταδεδομένων που συνοδεύουν κάθε αρχείο και γίνεται από κατάλληλες ρουτίνες που εκτελούνται τοπικά στον client, οι οποίες έχουν φορτωθεί κατά τη διαδικασία αίτησης για την αντίστοιχη σελίδα. Ο τελικός έλεγχος

πραγματοποιείται στην αντίπερα όχθη και αφορά την εξέταση του προς αποθήκευση αρχείου. Οι ρουτίνες που είναι υπεύθυνες για τη διαδικασία εκτελούνται στο μηχάνημα του server και τα πιθανά λάθη αναφέρονται πίσω στον client. Η διαδικασία επαναλαμβάνεται έως ότου το υποσύστημα αναφέρει επιτυχία με το αρχείο τελικά να αποθηκεύεται σε κατάλληλο φάκελο και την πληροφορία που το συνοδεύει να εισάγεται στην ψηφιακή διαλεκτική βάση.

Δοκιμάστηκε και ελέγχθηκε η ευρωστία του συστήματος της Βάσης Δεδομένων εισάγοντας δεκάδες εγγραφές.

Στον έλεγχο λάβαμε υπ' όψιν τις λειτουργικές προδιαγραφές και τις δυνατότητες των εργαλείων που χρησιμοποιήθηκαν. Η σχεδίαση είναι το πιο σημαντικό κομμάτι στην



Σχήμα 3: Βασική συσχέτιση των στοιχείων της βάσης δεδομένων

δημιουργία μιας νέας βάσης. Αρχικά οι πίνακες που δημιουργήθηκαν και σχεδιάστηκαν με τρόπο ώστε να μην περιέχεται επαναλαμβανόμενη πληροφορία αλλά και να μην χάνονται οι σχέσεις μεταξύ χαρακτηριστικών. Επίσης, σημαντικός είναι και ο χρόνος αναζήτησης και έτσι η βάση σχεδιάστηκε ώστε να βελτιστοποιηθεί όσο δυνατόν περισσότερο, σε συνδυασμό με τις ερωτήσεις που θα δέχεται η βάση.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
Πρόγραμμα για τη ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Έγινε διεξοδικός έλεγχος και βελτιστοποίηση της βάσης δεδομένων. Η βάση δεδομένων αποτελείται τελικά από επτά πίνακες ο οποίοι αποθηκεύουν όλη την πληροφορία που δίνεται από τους ερευνητές.

Ο πίνακας για τα αρχεία ήχου περιέχει ως χαρακτηριστικά:

- ID μοναδικό για κάθε αρχείο
- IDομιλητή
- Όνομα αρχείου
- IDηχογράφησης
- Τύπος επικοινωνίας
- Μονοπάτι (που βρίσκεται το αρχείο στον server)
- Ημερομηνία

Ο πίνακας για τους ομιλητές περιέχει ως χαρακτηριστικά:

- ID μοναδικό για κάθε ομιλητή
- Όνομα ομιλητή
- Επώνυμο ομιλητή
- Διαμέρισμα
- Νομός
- Ακριβής τοποθεσία
- Longitude
- Latitude
- Φύλο
- Ηλικία
- Ηλικιακή ομάδα
- Εκπαίδευση
- Επάγγελμα
- Ομάδα καταγωγής
- Καθεστώς Καταγωγής
- Δομή γλωσσικής κοινότητας
- Ύπαρξη προσφύγων
- Διάρκεια παραμονής στη Γλωσσική Κοινότητα
- Δομή Γειτονιάς

Ο πίνακας για τα αρχεία textgrid περιέχει ως χαρακτηριστικά:

- ID μοναδικό για κάθε textgrid
- IDαρχείου
- Όνομα textgrid

- ΙΔηχογράφησης
- Μονοπάτι (που βρίσκεται το textgrid στον server)

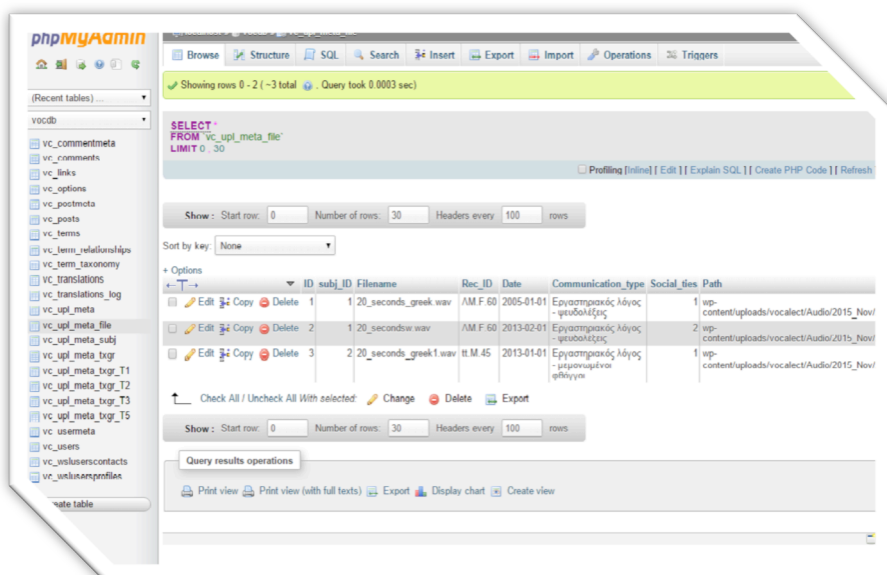
Ο πίνακας για τα tier 1/2 περιέχει ως χαρακτηριστικά:

- ΙΔαρχείου
- xmin
- xmax
- text

Ο πίνακας για τα tier 3/5 περιέχει ως χαρακτηριστικά:

- ΙΔαρχείου
- time
- mark

Η συσχέτιση όπως φαίνεται και στο Σχήμα 2, εκφράζει το πως τα δεδομένα ενώ αποθηκεύονται σε διαφορετικούς πίνακες, διατηρούν την σχέση την οποία έχουν μεταξύ τους. Για κάθε αρχείο ήχου που εισάγεται στη βάση, στις πληροφορίες του εισάγεται και ο μοναδικός αριθμός του ομιλητή. Αντίστοιχα, για κάθε textgrid και tier εισχωρήσεις στις πληροφορίες τους εισάγεται και ο μοναδικός αριθμός του αρχείου ήχου. Με τον τρόπο αυτό διευκολύνεται και η αναζήτηση μεταξύ πινάκων μειώνοντας σημαντικά τον όγκο πληροφορίας που πρέπει ο υπολογιστής να διαβάσει για να βγάλει αποτελέσματα.



ID	subj_ID	Filename	Rec_ID	Date	Communication_type	Social_files	Path
1	1	20_seconds_greek.wav	AM F.60	2005-01-01	Εργαστηριακός λόγος - ψευδολέξεις	1	wp-content/uploads/vocalect/Audio/2015_Nov/
2	1	20_seconds.wav	AM F.60	2013-02-01	Εργαστηριακός λόγος - ψευδολέξεις	2	wp-content/uploads/vocalect/Audio/2015_Nov/
3	2	20_seconds_greek1.wav	tt.M.45	2013-01-01	Εργαστηριακός λόγος - μεμονωμένα φθόγγοι	1	wp-content/uploads/vocalect/Audio/2015_Nov/

Σχήμα 4: Παράδειγμα ενός πίνακα της βάσης



4. Συμπεράσματα

Κατά την διάρκεια του ελέγχου εντοπίστηκαν μικρά προβλήματα τόσο στο σχεδιασμό όσο και στην ανάπτυξη για τα οποία και προτάθηκαν διορθωτικές παρεμβάσεις οι οποίες και υλοποιήθηκαν.